

Wiley Series in Survey Methodology

Improving Surveys with Paradata

Analytic Uses of Process Information



Edited by

Frauke Kreuter

WILEY

IMPROVING SURVEYS WITH PARADATA

WILEY SERIES IN SURVEY METHODOLOGY

Established in Part by WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors: *Mick P. Couper, Graham Kalton, J. N. K. Rao, Norbert Schwarz,
Christopher Skinner*

Editor Emeritus: *Robert M. Groves*

A complete list of the titles in this series appears at the end of this volume.

IMPROVING SURVEYS WITH PARADATA

Analytic Uses of Process Information

Edited by

FRAUKE KREUTER

Joint Program in Survey Methodology, University of Maryland
Institute for Employment Research, Nuremberg
Ludwig Maximilian University, Munich

WILEY

Cover Design: John Wiley & Sons, Inc.
Cover Illustration: Courtesy of Frauke Kreuter

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Improving surveys with paradata: analytic uses of process information / [edited by] Frauke Kreuter, University of Maryland, College Park, Maryland, Institute for Employment Research, Nuremberg, Ludwig-Maximilians-University, Munich.

pages cm

Includes bibliographical references and index.

ISBN 978-0-470-90541-8 (cloth)

1. Surveys—Statistical methods. 2. Social surveys—Statistical methods. 3. Social sciences—Research—Statistical methods. I. Kreuter, Frauke.

HA31.2.147 2013

001.4'33—dc23

2013000328

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE	xv
CONTRIBUTORS	xix
ACRONYMS	xxi
1 IMPROVING SURVEYS WITH PARADATA: INTRODUCTION	1
<i>Frauke Kreuter</i>	
1.1 Introduction	1
1.2 Paradata and Metadata	2
1.3 Auxiliary Data and Paradata	3
1.4 Paradata in the Total Survey Error Framework	4
1.5 Paradata in Survey Production	6
1.6 Special Challenges in the Collection and Use of Paradata	6
1.6.1 Mode-Specific Paradata	6
1.6.2 Complex Structure	7
1.6.3 Quality of Paradata	7
1.7 Future of Paradata	8
References	9

PART I PARADATA AND SURVEY ERRORS

2 PARADATA FOR NONRESPONSE ERROR INVESTIGATION 13*Frauke Kreuter and Kristen Olson*

2.1	Introduction	13
2.2	Sources and Nature of Paradata for Nonresponse Error Investigation	14
2.2.1	Call History Data	14
2.2.2	Interviewer Observations	17
2.2.3	Measures of the Interviewer–Householder Interaction	19
2.3	Nonresponse Rates and Nonresponse Bias	20
2.3.1	Studying Nonresponse with Paradata	22
2.3.2	Call Records	22
2.3.3	Interviewer Observations	27
2.3.4	Observations of Interviewer–Householder Interactions	29
2.4	Paradata and Responsive Designs	30
2.5	Paradata and Nonresponse Adjustment	31
2.6	Issues in Practice	32
2.7	Summary and Take Home Messages	34
	References	34

3 COLLECTING PARADATA FOR MEASUREMENT ERROR EVALUATIONS 43*Kristen Olson and Bryan Parkhurst*

3.1	Introduction	43
3.2	Paradata and Measurement Error	44
3.3	Types of Paradata	47
3.3.1	Time Stamps	47
3.3.2	Keystrokes	49
3.3.3	Mouse Clicks	50
3.3.4	Behavior Codes	51
3.3.5	Vocal Characteristics	53
3.3.6	Interviewer Evaluations	54
3.4	Differences in Paradata by Modes	56
3.4.1	In-Person Surveys	56
3.4.2	Telephone Surveys	57
3.4.3	Web Surveys	60

3.5	Turning Paradata into Datasets	61
3.5.1	Paradata as Text Files	61
3.5.2	Paradata as Sound Files	64
3.5.3	Paradata as Variables	64
3.6	Summary	65
	Funding Note	65
	References	65

4 ANALYZING PARADATA TO INVESTIGATE MEASUREMENT ERROR 73

Ting Yan and Kristen Olson

4.1	Introduction	73
4.2	Review of Empirical Literature on the Use of Paradata for Measurement Error Investigation	73
4.2.1	Using Paradata to Understand the Question–Answering Process	74
4.2.2	Using Paradata to Investigate Usability Issues in CAI Systems	75
4.2.3	Reduction of Measurement Error	75
4.2.4	Adjusting for Measurement Error	76
4.3	Analyzing Paradata	76
4.3.1	Units of Analysis	77
4.3.2	Data Management	78
4.3.3	Other Auxiliary Variables	82
4.3.4	Modeling Decisions	82
4.4	Four Empirical Examples	83
4.4.1	Draisma and Dijkstra (2004)—Item Level: Paradata as an Independent Variable	83
4.4.2	Malhotra (2008)—Survey Level: Paradata as an Independent Variable	86
4.4.3	Yan and Tourangeau (2008)—Item Level: Paradata as a Dependent Variable	87
4.4.4	Lenzner, Kaczmirek, and Lenzner (2010)—Survey Level: Paradata as a Dependent Variable	88
4.5	Cautions	89
4.6	Concluding Remarks	90
	References	90

5 PARADATA FOR COVERAGE RESEARCH	97
<i>Stephanie Eckman</i>	
5.1 Introduction	97
5.2 Housing Unit Frames	101
5.2.1 Postal Delivery Databases	101
5.2.2 Housing Unit Listing	104
5.2.3 Random Route Sampling	106
5.2.4 Missed Unit Procedures	108
5.3 Telephone Number Frames	109
5.4 Household Rosters	111
5.5 Population Registers	113
5.6 Subpopulation Frames	113
5.7 Web Surveys	114
5.8 Conclusion	115
Acknowledgments	115
References	116

PART II PARADATA IN SURVEY PRODUCTION

6 DESIGN AND MANAGEMENT STRATEGIES FOR PARADATA-DRIVEN RESPONSIVE DESIGN: ILLUSTRATIONS FROM THE 2006–2010 NATIONAL SURVEY OF FAMILY GROWTH	123
<i>Nicole G. Kirgis and James M. Lepkowski</i>	
6.1 Introduction	123
6.2 From Repeated Cross-Section to Continuous Design	124
6.3 Paradata Design	129
6.4 Key Design Change 1: A New Employment Model	134
6.5 Key Design Change 2: Field Efficient Sample Design	135
6.6 Key Design Change 3: Replicate Sample Design	137
6.7 Key Design Change 4: Responsive Design Sampling of Nonrespondents in a Second Phase	139
6.8 Key Design Change 5: Active Responsive Design Interventions	140
6.9 Concluding Remarks	141
References	143

7 USING PARADATA-DRIVEN MODELS TO IMPROVE CONTACT RATES IN TELEPHONE AND FACE-TO-FACE SURVEYS 145

James Wagner

7.1	Introduction	145
7.2	Background	146
7.3	The Survey Setting	148
7.4	Experiments: Data and Methods	149
7.4.1	Call Windows	150
7.4.2	The Data	151
7.4.3	The Models	154
7.4.4	Procedure: Telephone Survey	157
7.4.5	Procedure: Face-to-Face Survey	160
7.5	Experiments: Results	161
7.5.1	Telephone Survey	161
7.5.2	Face-to-Face Survey	165
7.6	Discussion	166
	References	169

8 USING PARADATA TO STUDY RESPONSE TO WITHIN-SURVEY REQUESTS 171

Joseph W. Sakshaug

8.1	Introduction	171
8.2	Consent to Link Survey and Administrative Records	175
8.2.1	Modeling Linkage Consent Using Paradata: Example from the Health and Retirement Study	176
8.2.2	Using Paradata for Intervention	178
8.3	Consent to Collect Biomeasures in Population-Based Surveys	178
8.3.1	Modeling Biomeasure Consent Using Paradata: Example from the Health and Retirement Study	179
8.3.2	Using Paradata for Intervention	180
8.4	Switching Data Collection Modes	180
8.4.1	Predicting Mode Switch Response Using Paradata: Example from a Survey of University Alumni	181
8.4.2	Using Paradata for Intervention	182
8.5	Income Item Nonresponse and Quality of Income Reports	183
8.5.1	Studying Income Item Nonresponse and Quality of Income Reports Using Paradata: Examples from the Health and Retirement Study	184
8.5.2	Using Paradata for Intervention	186

8.6	Summary	186
	Acknowledgments	187
	References	187
9	MANAGING DATA QUALITY INDICATORS WITH PARADATA BASED STATISTICAL QUALITY CONTROL TOOLS: THE KEYS TO SURVEY PERFORMANCE	191
	<i>Matt Jans, Robyn Sirkis, and David Morgan</i>	
9.1	Introduction	191
9.2	Defining and Choosing Key Performance Indicators (KPIs)	193
9.3	KPI Displays and the Enduring Insight of Walter Shewhart	200
9.3.1	Understanding a Process: Impediments to Clear Quality Control Steps	205
9.3.2	Rules for Finding Special Cause Variation in a Control Chart	207
9.4	Implementation Steps for Survey Analytic Quality Control with Paradata Control Charts	210
9.5	Demonstrating a Method for Improving Measurement Process Quality Indicators	213
9.6	Reflections on SPC, Visual Data Displays, and Challenges to Quality Control and Assurance with Survey Analytics	220
9.7	Some Advice on Using Charts	221
	Appendix	224
	Acknowledgments	226
	References	226
10	PARADATA AS INPUT TO MONITORING REPRESENTATIVENESS AND MEASUREMENT PROFILES: A CASE STUDY OF THE DUTCH LABOUR FORCE SURVEY	231
	<i>Barry Schouten and Melania Calinescu</i>	
10.1	Introduction	231
10.2	Measurement Profiles	233
10.2.1	What are Measurement Profiles?	233
10.2.2	Measurement Profiles in the Labour Force Survey	234
10.3	Tools for Monitoring Nonresponse and Measurement Profiles	236
10.4	Monitoring and Improving Response: A Demonstration Using the LFS	242
10.4.1	The Choice of Design Features in Monitoring Nonresponse and Measurement Profiles	242
10.4.2	The Choice of Population Subgroups in Monitoring	243

10.4.3 Partial R-indicators for the Sequential and Cumulative Analysis	245
10.4.4 Adaptive Survey Designs	250
10.5 Including Paradata Observations on Households and Persons	253
10.6 General Discussion	255
10.7 Take Home Messages	256
Acknowledgments	257
References	257

PART III SPECIAL CHALLENGES

11 PARADATA IN WEB SURVEYS 261

Mario Callegaro

11.1 Survey Data Types	261
11.2 Collection of Paradata	262
11.3 Typology of Paradata in Web Surveys	262
11.3.1 Uses of Paradata: Device Type	264
11.3.2 Uses of Paradata: Questionnaire Navigation	267
11.4 Using Paradata to Change the Survey in Real Time: Adaptive Scripting	271
11.5 Paradata in Online Panels	272
11.6 Software to Collect Paradata	272
11.7 Analysis of Paradata: Levels of Aggregation	273
11.8 Privacy and Ethical Issues in Collecting Web Survey Paradata	274
11.9 Summary and Conclusions on Paradata in Web Surveys	275
References	276

12 MODELING CALL RECORD DATA: EXAMPLES FROM CROSS-SECTIONAL AND LONGITUDINAL SURVEYS 281

Gabriele B. Durrant, Julia D'Arrigo, and Gerrit Müller

12.1 Introduction	281
12.2 Call Record Data	283
12.3 Modeling Approaches	285
12.3.1 Analysis Approaches and the Use of Multilevel Modeling	285
12.3.2 Specifications of Multilevel Discrete-Time Event History Models for the Analysis of Call Record Data	287
12.3.3 Modeling Strategy and Estimation of Models	292

12.4	Illustration of Call Record Data Analysis Using Two Example Datasets	293
12.4.1	Analyzing Call Outcomes in the PASS Longitudinal Survey	293
12.4.2	Analyzing Call Outcomes in the UK Census Nonresponse Link Study	296
12.5	Summary	304
	Acknowledgments	304
	References	304
13	BAYESIAN PENALIZED SPLINE MODELS FOR STATISTICAL PROCESS MONITORING OF SURVEY PARADATA QUALITY INDICATORS	309
	<i>Joseph L. Schafer</i>	
13.1	Introduction	309
13.1.1	Processes Under Control and Out of Control	309
13.1.2	Motivating Example	311
13.1.3	Looking Ahead	313
13.2	Overview of Splines	313
13.2.1	Definition	313
13.2.2	Basis Functions	315
13.2.3	Parameters of Interest	316
13.2.4	Branching Splines	318
13.2.5	Knot Density and Roughness	319
13.3	Penalized Splines as Linear Mixed Models	321
13.3.1	Model Formulation	321
13.3.2	Estimating Parameters	322
13.3.3	Estimating the Function	323
13.3.4	Difficulties with Likelihood Inference	324
13.4	Bayesian Methods	325
13.4.1	Bayesian Inference for the Smoothing Parameter	325
13.4.2	Bayesian Intervals and Predictions	326
13.5	Extensions	329
	Appendix	330
A.1	Maximum-Likelihood Estimation	330
A.2	Posterior Simulation	333
A.3	Bayesian Inference About the Mean Function	335
A.4	Disclaimer	337
	References	337

14 THE QUALITY OF PARADATA: A LITERATURE REVIEW	339
<i>Brady T. West and Jennifer Sinibaldi</i>	
14.1 Introduction	339
14.2 Existing Studies Examining the Quality of Paradata	340
14.2.1 Computer-Generated Process Data	340
14.2.2 Interviewer-Recorded Call Record Data	343
14.2.3 Interviewer Observations	349
14.3 Possible Mechanisms Leading to Error in Paradata	352
14.3.1 Computer-Generated Process Data	352
14.3.2 Interviewer Observations	353
14.4 Take Home Messages	356
References	356
 15 THE EFFECTS OF ERRORS IN PARADATA ON WEIGHTING CLASS ADJUSTMENTS: A SIMULATION STUDY	 361
<i>Brady T. West</i>	
15.1 Introduction	361
15.2 Design of Simulation Studies	364
15.2.1 Simulation Parameters	364
15.2.2 Alternative Estimators	368
15.3 Simulation Results	370
15.3.1 Scenario 1: D has a Positive Relationship with Y and a Negative Relationship with R , FNRs are Greater Than FPRs for Both Respondents and Nonrespondents, FNRs are Equal for Respondents and Nonrespondents, and FPRs are Equal for Respondents and Nonrespondents	370
15.3.2 Scenario 2: D has a Positive Relationship with Both Y and R , FNRs are Greater Than FPRs for Both Respondents and Nonrespondents, FNRs are Equal for Respondents and Nonrespondents, and FPRs are Equal for Respondents and Nonrespondents	373
15.3.3 Scenario 3: D has a Positive Relationship with Y and a Negative Relationship with R , FNRs are Greater Than FPRs for Both Respondents and Nonrespondents, FNRs for Nonrespondents are Greater Than Those for Respondents, and FPRs for Nonrespondents are Greater Than Those for Respondents	376

15.3.4 Scenario 4: D has a Positive Relationship with Both Y and R , FNRs are Greater Than FPRs for Both Respondents and Nonrespondents, FNRs for Nonrespondents are Greater Than Those for Respondents, and FPRs for Nonrespondents are Greater Than Those for Respondents	378
15.3.5 Additional Scenarios	379
15.4 Take Home Messages	383
15.5 Future Research	385
References	387
INDEX	389

PREFACE

Newspapers and blogs are now filled with discussions about “big data,” massive amounts of largely unstructured data generated by behavior that is electronically recorded. “Big data” was the central theme at the 2012 meeting of the World Economic Forum and the U.S. Government issued a Big Data Research and Development Initiative the same year. The American Statistical Association has also made the topic a theme for the 2012 and 2013 Joint Statistical Meetings.

Paradata are a key feature of the “big data” revolution for survey researchers and survey methodologists. The survey world is peppered with process data, such as electronic records of contact attempts and automatically captured mouse movements that respondents produce when answering web surveys. While not all of these data sets are massive in the usual sense of “big data,” they are often highly unstructured, and it is not always clear to those collecting the data which pieces are relevant, and how they should be analyzed. In many instances it is not even obvious which data are generated.

Recently Axel Yorder, the CEO of the company Webtrends, pointed out that just as “Gold requires mining and processing before it finds its way into our jewelry, electronics, and even the Fort Knox vault [...] data requires collection, mining and, finally, analysis before we can realize its true value for businesses, governments, and individuals alike.”¹ The same can be said for paradata. Paradata are data generated in the process of conducting a survey. As such, they have the potential to shed light on the survey process itself, and with proper “mining” they can point to errors and breakdowns in the process of data collection. If captured and analyzed immediately paradata can assist

¹http://news.cnet.com/8301-1001_3-57434736-92/big-data-is-worth-nothing-without-big-science/

with efficiency during data collection field period. After data collection ends, paradata that capture measurement errors can be modeled alongside the substantive data to increase the precision of resulting estimates. Paradata collected for respondents and nonrespondents alike can be useful for nonresponse adjustment. As discussed in several chapters in this volume, paradata can lead to efficiency gains and cost savings in survey data production. This has been demonstrated in the U.S. National Survey of Family Growth conducted by the University of Michigan and the National Center for Health Statistics.

However, just as for big data in general, many questions remain about how to turn paradata into gold. Different survey modes allow for the collection of different types of paradata, and depending on the production environment, paradata may be instantaneously available. Fast-changing data collection technology will likely open doors to real-time capture and analysis of even more paradata in ways we cannot currently imagine. Nevertheless some general principles regarding the logic, design, and use of paradata will not change, and this book discusses these principles. Much work in this area is done within survey research agencies and often does not find its way into print, thus this book also serves as a vehicle to share current developments in paradata research and use.

This book came to life during a conference sponsored by the Institute for Employment Research in Germany, November of 2011 when most of the chapter authors participated in a discussion about it. The goal was to write a book that goes into more detail than published papers on the topic. Because this research area is relatively new we saw the need to collect information that is otherwise not easily accessible and to give practitioners a good starting point for their own work with paradata. The team of authors decided to use a common framework and standardized notation as much as possible. We tried to minimize overlap across the chapters without hampering the possibility for each chapter to be read on its own. We hope the result will satisfy the needs of researchers starting to use paradata as well as those who are already experienced. We also hope it will inspire readers to expand the use of paradata to improve survey data quality and survey processes. As we strive to update our knowledge on behalf of all authors, I ask you to tell us about your successes and failures in dealing with paradata.

We dedicate this volume to Mick Couper and Robert Groves. Mick Couper coined the term “paradata” in a presentation at the 1998 Joint Statistical Meeting in Dallas where he discussed the potential of paradata to reduce measurement error. For his vision regarding paradata he was awarded the American Association for Public Opinion Research’s Warren J. Mitofsky Innovators Award in 2008. As the director of the University of Michigan Survey Research Center and later as Director of the U.S. Census Bureau, Robert Groves implemented new ideas on the use of paradata to address nonresponse, showing the breadth of applications paradata have to survey errors and operational challenges. After a research seminar in the Joint Program in Survey Methodology on this topic, I remember him saying: “You should write a book on paradata!” Both Mick and Bob have been fantastic teachers and mentors for most of the chapter authors and outstanding colleagues to all. Their perspectives on Survey

Methodology and the Total Survey Error Framework are guiding principles visible in each of the chapters.

I personally also want to thank Rainer Schnell for exposing me to paradata before they were named as such. As part of the German DEFECT project that he led, we walked through numerous villages and cities in Germany to collect addresses. In this process we took pictures of street segments and recorded, on the first generation of handheld devices, observations and judgments about the selected housing units. Elizabeth Coutts, my dear friend and colleague in this project, died on August 5, 2009, but her ingenious contributions to the process of collecting these paradata will never be forgotten.

We are very grateful to Paul Biemer, Lars Lyberg and Fritz Scheuren for actively pushing the paradata research agenda forward and for making important contributions by putting paradata into the context of statistical process control and the larger meta-data initiatives. This book benefitted from discussions at the International Workshop on Household Survey Nonresponse and the International Total Survey Error Workshop and we are in debt to all of the researchers who shared their work and ideas at these venues over the years. In particular, we thank Nancy Bates, James Dahlhamer, Mirta Galesic, Barbara O'Hare, Rachel Horwitz, François Laflamme, Lars Lyberg, Andrew Mercer Peter Miller and Stanley Presser for comments on parts of this book. Our thanks also goes to Ulrich Kohler for creating the cover page graph.

The material presented here provided the basis for several short courses taught during the Joint Statistical Meeting of the American Statistical Association, continuing education efforts of the U.S. Census Bureau, the Royal Statistical Society, and the European Social Survey. The feedback I received from course participants helped to improve this book, but remaining errors are entirely ours.

On the practical side, this book would not have found its way into print without our LaTeX wizard Alexandra Birg, the constant pushing of everybody involved at Wiley, and the support from the Joint Program in Survey Methodology in Maryland, the Institute for Employment Research in Nuremberg, and the Department of Statistics at the Ludwig Maximilian University in Munich. We thank you all.

FRAUKE KREUTER

*Washington D.C.
September, 2012*

CONTRIBUTORS

MELANIA CALINESCU, VU University Amsterdam, NL

MARIO CALLEGARO, Google London, UK

JULIA D'ARRIGO, Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, Southampton, UK

GABRIELE B. DURRANT, Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, Southampton, UK

STEPHANIE ECKMAN, Institute for Employment Research (IAB), Nuremberg, Germany

MATT JANS, University of California Los Angeles, Los Angeles, California, USA

NICOLE G. KIRGIS, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

FRAUKE KREUTER, Institute for Employment Research (IAB), Nuremberg, Germany; University of Maryland, College Park, Maryland, USA; Ludwig Maximilian University, Munich, Germany

JAMES M. LEPKOWSKI, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

DAVID MORGAN, U.S. Census Bureau, Washington, DC, USA

GERRIT MÜLLER, Institute for Employment Research (IAB), Nuremberg, Germany

KRISTEN OLSON, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

BRYAN PARKHURST, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

JOSEPH W. SAKSHAUG, Institute for Employment Research (IAB), Nuremberg, Germany

JOSEPH L. SCHAFER, Center for Statistical Research and Methodology, U.S. Census Bureau, Washington, DC, USA

BARRY SCHOUTEN, Statistics Netherlands, Den Haag and University of Utrecht, NL

JENNIFER SINIBALDI, Institute for Employment Research (IAB), Nuremberg, Germany

ROBYN SIRKIS, U.S. Census Bureau, Washington DC, USA

JAMES WAGNER, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan

BRADY T. WEST, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

TING YAN, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

ACRONYMS

AAPOR	American Association for Public Opinion Research
ACASI	Audio Computer-Assisted Self-Interview
ACS	The American Community Survey
AHEAD	Assets and Health Dynamics Among the Oldest Old
ANES	American National Election Studies
BCS	British Crime Survey
CAI	Computer-Assisted Interviewing
CAPI	Computer-Assisted Personal Interviews
CARI	Computer-Assisted Recording of Interviews
CASRO	Council of American Survey Research Organizations
CATI	Computer-Assisted Telephone Interviews
CE	Consumer Expenditure Interview Survey
CHI	Contact History Instrument
CHUM	Check for Housing Unit Missed
CPS	Current Population Survey
CSP	Client-side Paradata
ESOMAR	European Society for Opinion and Market Research
ESS	European Social Survey
FRS	Family Resources Survey
GSS	General Social Survey
HINTS	Health Information National Trends Study
HRS	Health and Retirement Study
IAB	Institute for Employment Research
IVR	Interactive Voice Response System
KPI	Key Performance Indicators

LAFANS	Los Angeles Family and Neighborhood Study
LCL	Lower Control Limits
LFS	Labour Force Survey
LISS	Dutch Longitudinal Internet Studies for the Social Sciences
LMU	Ludwig Maximilian University Munich
NCHS	National Center for Health Statistics
NHANES	National Health and Nutrition Examination Survey
NHEFS	The NHANES Epidemiologic Follow-up Study
NHIS	National Health Interview Survey
NSDUH	National Survey of Drug Use and Health
NSFG	National Survey of Family Growth
NSHAP	National Social Life, Health, and Aging Project
NSR	Non-self Representing
OMB	Office of Management and Budget
PASS	Panel Study of Labour Market and Social Security
PDA	Personal Digital Assistant
PSU	Primary Sampling Units
RDD	Random Digit Dial
RECS	Residential Energy Consumption Survey
RMSE	Root Mean Squared Error
RO	Regional Office
SCA	Survey of Consumer Attitudes
SCF	Survey of Consumer Finances
SHS	Survey of Household Spending
SPC	Statistical Process Control
SQC	Statistical Quality Control
SR	Self-Representing Areas
UCL	Upper Control Limits
UCSP	Universal Client Side Paradata

CHAPTER 1

IMPROVING SURVEYS WITH PARADATA: INTRODUCTION

FRAUKE KREUTER

University of Maryland and IAB/LMU

1.1 INTRODUCTION

Good quality survey data are hard to come by. Errors in creating proper representation of the population and errors in measurement can threaten the final survey estimates. Survey methodologists work to improve survey questions, data entry interfaces, frame coverage, sampling procedures, respondent recruitment, data collection, data editing, weighting adjustment procedures, and many other elements in the survey data production process to reduce or prevent errors. To study errors associated with different steps in the survey production process, researchers have used experiments, benchmark data, or simulation techniques as well as more qualitative methods, such as cognitive interviewing or focus groups. The analytic use of paradata now offers an additional tool in the survey researcher's tool box to study survey errors and survey costs. The production of survey data is a process that involves many actors, who often must make real time decisions informed by observations from the ongoing data collection process. What observations are used for decision making and how those decisions are made are currently often outside the researchers' direct control. A few examples: *Address listers* walk or drive around neighborhoods, making decisions about the inclusion or exclusion of certain housing units based on their perceptions of the housing and neighborhood characteristics. *Field managers* use personal experience and subjective judgment to instruct interviewers to intensify or reduce their efforts on specific cases. *Interviewers* approach households and conduct interviews in idiosyncratic ways; doing so they might use observations about the sampled households to tailor their approaches. *Respondents* answer survey questions in settings unknown to the researcher but which affect their responses; they might be interrupted when answering a web survey, or other family members might join the conversation the respondent is having with the interviewer. Wouldn't we like to have a bird's eye

view to know what was going on in each of these situations? What information does a particularly successful field manager use when assigning cases? Which strategy do particularly successful interviewers use when recruiting respondents? What struggles does a respondent have when answering a survey question? With this knowledge we could tweak the data collection process or analyze the data differently. Of course, we could ask each and every one of these actors involved, but aside from the costs of doing so, much of what is going on is not necessarily a conscious process, and might not be stored in a way that it can be easily recalled (Tourangeau et al., 2000).

At the turn of the twenty-first century much of this process information became available, generated as a by-product of computer-assisted data collection. Mick Couper referred to these data as “paradata” in a presentation at the Joint Statistical Meeting in Dallas (Couper, 1998). Respondents in web surveys leave electronic traces as they answer survey questions, captured through their keystrokes and mouse clicks. In telephone surveys, automated call scheduling systems record the date and time of every call. In face-to-face surveys, interviewers’ keystrokes are easily captured alongside the interview and so are audio or even video recordings of the respondent–interviewer interactions. Each of these is an example of paradata available through the computerized survey software.

Some survey organizations have collected such information about the data collection process long before the rise of computer-assisted interviewing and the invention of the word paradata. However, a rapid growth in the collection and use of paradata can be seen in recent years (Scheuren, 2005). It is facilitated first, by the increase in computer-aided data collection around the world, second, by the increasing ease with which paradata are accessed, and third, by an increasing interest among survey sponsors in process quality and the quantification of process errors. Thus, while process quality and paradata are not new, a more structured approach in choosing, measuring, and analyzing key process variables is indeed a recent development (Couper and Lyberg, 2005). This book takes this structured approach and provides a summary of what we know to date about how paradata should be collected and used to improve survey quality, in addition to introducing new research results.

The chapters in the first part of this book review the current use of paradata and make general suggestions about paradata design principles. The second section includes several case studies for the use of paradata in survey production, either concurrently or through post hoc evaluations of production features. Chapters in the last section discuss challenges involved in the collection and use of paradata, including the collection of paradata in web surveys.

Before reading the individual book chapters, it is helpful to discuss some common definitions and to gain an overview of the framework that shaped the structure of this book and the write-up of the individual chapters.

1.2 PARADATA AND METADATA

There is no standard definition in the literature of what constitutes paradata. Papers discussing paradata vary in terminology from one to another (Scheuren, 2000; Couper

and Lyberg, 2005; Scheuren, 2005; O'Reilly, 2009), but for the purpose of the book we define paradata as additional data that can be captured during the process of producing a survey statistic. Those data can be captured at all stages of the survey process and with very different granularities. For example, response times can be captured for sets of questions, one question and answer sequence, or just for the answer process itself.

There is some debate in the literature over how paradata differ from metadata. Metadata are often described as data about data, which seems to greatly overlap with our working definition of paradata. Let us step back for a moment and consider an analogy to digital photography which may make the paradata–metadata distinction clearer. Digital information such as the time and day a picture was taken is often automatically added by cameras to the file. Similarly, the lens and exposure time and other settings that were used can be added to the file by the photographer. In the IT setting, this information is called metadata or data about data.

Paradata are instead data about the process of generating the final product, the photograph or the survey dataset. In the photography example, the analogy to paradata would be data that capture which lenses were tried before the final picture was taken, information about different angles the photographer tried before producing the final shot, and the words she called out before she was able to make the subject smile.

In the digital world, metadata have been a common concept for quite a while. In the social sciences, the interest in metadata is newer but heavily promoted through efforts like the Data Documentation Initiative or DDI (<http://www.ddialliance.org/>), which is a collaboration between European and U.S. researchers to develop standards for social science data documentation. Metadata are the core of this documentation and can be seen as macro-level information about survey data; examples are information about the sampling frame, sampling methods, variable labels, value labels, percentage of missing data for a particular variable, or the question text in all languages used for the survey. Metadata allow users to understand the structure of a dataset and can inform analysis decisions.

Paradata capture information about the data collection process on a more micro-level. Some of this information forms metadata if aggregated, for example, the response rate for a survey (a piece of metadata) is an aggregated value across the case-level final result codes. Or, using the examples given above, time measurements could be aggregated up to become metadata. Paradata that capture the minutes needed to interview each respondent or even the seconds it took to administer a single question within the survey would become the metadata information on the average time it took to administer the survey.

1.3 AUXILIARY DATA AND PARADATA

Paradata are not the only source of additional data used in survey research to enrich final datasets and estimates. Researchers also use what they call ‘auxiliary data’, but the definition of this term has not quite been settled upon. The keyword auxiliary data has been used to encompass all data outside of the actual survey data itself, which

would make all paradata also auxiliary data. Also contained under auxiliary data are variables from the sampling frame and data that can be linked from other sources. The other sources are often from the Census or American Community Survey, or other government agencies and private data collectors. They are typically available on a higher aggregate level than the individual sampling unit, for example, city blocks or block groups or tracts used for Census reports or voting registries. Unlike paradata, they tend to be fixed for a given sampling unit and available outside of the actual data collection process. A typical example would be the proportion of minority households in a given neighborhood or block according to the last Census.

Paradata, as we define them here, are not available prior to data collection but generated within, and they can change over the course of the data collection. A good example is interviewer experience within the survey. If the sequence of contact attempts is analyzed and interviewer experience is added to the model, it would form a time varying covariate, for the experience changes with every case the interviewer worked on. Data on interviewer demographic characteristics are not always easily classified as either paradata or auxiliary variables. Technically, those data collected outside the survey are auxiliary data that can be merged to the survey data. However, if we think of the process of recruiting respondents, there might be changes throughout the survey in which cases are re-assigned to different interviewers, so the characteristics associated with the case (which include interviewer characteristics) might change because the interviewer changes.

A large set of different auxiliary data sources available for survey researchers was discussed at the 2011 International Nonresponse Workshop (Smith, 2011), where paradata were seen as one of many sources of auxiliary data. In the context of this book, we focus on paradata, because compared to other auxiliary data sources, their collection and use is more likely under the control of survey practitioners.

1.4 PARADATA IN THE TOTAL SURVEY ERROR FRAMEWORK

Paradata can help researchers understand and improve survey data. When we think about the quality of survey data, or more specifically a resulting survey statistic, the Total Survey Error Framework is a helpful tool. Groves et al. (2004) visualized the data collection process in two strands, one reflecting steps necessary for representation, the other steps necessary for measurement (see Figure 1.1). Each of the steps carries the risk of errors. When creating a sampling frame, there is a chance to miss some members of the population or to include those that do not belong, both of which can lead to coverage error. Sampling errors refer to the imprecision resulting from surveying only a sample instead of the population, usually reflected in standard error estimates. If selected cases refuse to participate in the survey, methodologists talk about nonresponse error, and any failure to adjust properly for such selection processes will result in adjustment error. On the measurement side, if questions fail to reflect the underlying concepts of interest, they suffer from low validity. Even when questions perfectly measure what is of interest to the researcher, failures can occur in the response process, leading to measurement error. Survey production often includes

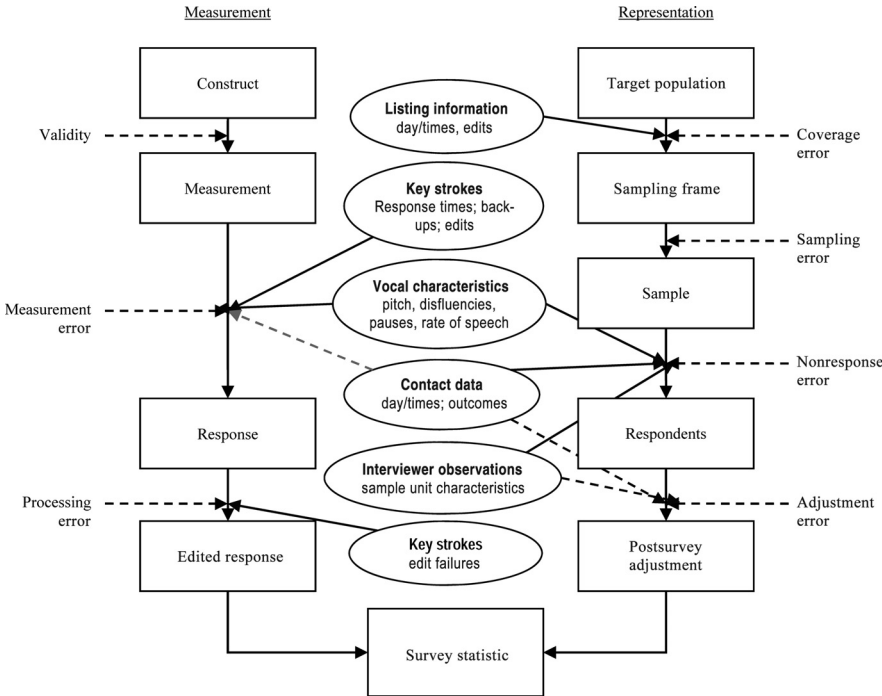


FIGURE 1.1 Survey process and process data collected to inform each of the total survey error components (graph modified from Groves et al. (2004), and expanded from Kreuter and Casas-Cordero (2010)). Solid lines mark paradata collected at a particular step; dashed lines (leaving the ovals) indicate that paradata are used to evaluate errors at the particular step, even though they are not collected during this step.

a phase of editing involving important consistency checks, and things can go wrong at this step too. Paradata can inform researchers about such errors that can happen along the way. In some instances, they can point to problems that can be solved during data collection; in other instances, they capture the information needed to model the errors alongside the actual survey data. Figure 1.1 depicts, within the survey data production process and the associated survey errors, some examples of paradata that are either collected at the respective steps (marked with a solid arrow) or used to evaluate a given error source (marked with a dashed arrow).

The chapters in the first section of this book are designed to introduce paradata within the Total Survey Error Framework. So far, paradata related to nonresponse are featured most prominently in the survey literature. The findings in these areas are discussed in detail by Frauke Kreuter, Kristen Olson, Bryan Packhurst, and Ting Yan. Paradata which inform us about coverage error are of increasing interest in a world with multiple frame creation methods, and are discussed by Stephanie Eckman. Unfortunately, the literature on paradata to inform data processing and related errors is very sparse so far. Thus, there is no chapter addressing this error source, though the general logic of designing and capturing paradata for the other error sources applies

here too. Sampling errors and adjustment errors have been widely discussed in the literature, but as with coverage error, much less is done in terms of evaluating the process of sampling or adjustment through paradata. The same holds for the issue of validity, though one could imagine process information about questionnaire creation.

1.5 PARADATA IN SURVEY PRODUCTION

Paradata are not just used to evaluate survey errors after data collection is done. In some instances, paradata are available during data collection and can be used to monitor and inform the collection process in (almost) real time. Survey methodologists have started to explore using paradata to guide data collection procedures, a process called responsive or adaptive design. The chapter by Nicole Kirgis and James Lepkowski shares experiences using such an approach in the National Survey of Family Growth. Similar in spirit is the use of paradata to predict responses to within-survey requests, suggested by Joseph Sakshaug in Chapter 8. James Wagner reports paradata-driven experiments he carried out to try to increase response rates in both telephone and face-to-face surveys.

In order to monitor incoming data and to make useful design decisions, the field needs tools that display and summarize the large amount of incoming information. Some survey organizations, including the U.S. Census Bureau, have applied theories and methods from the quality control literature to their survey processes. These efforts are summarized in Chapter 9 by Matt Jans, Roby Sirkis, and David Morgan. Statistics Netherlands is now heavily engaged in using metrics to monitor representativeness in respondent composition as Barry Schouten and Melania Calinescu explain in Chapter 10.

1.6 SPECIAL CHALLENGES IN THE COLLECTION AND USE OF PARADATA

Despite the promise and hope of paradata, this new data source does present several challenges with which researchers are grappling. A few are mentioned here and are discussed in detail in the respective chapters. Others can only be touched on in this book, but are equally important.

1.6.1 Mode-Specific Paradata

The type of paradata that can be collected in a given survey or that is already available for a particular survey varies with the survey mode. Most examples discussed throughout this edited volume come from face-to-face surveys, and some from telephone surveys. Most self-administered surveys involve no interviewers and thus are stripped of one important vehicle for paradata collection. This is, however, not to say that self-administered surveys cannot be paradata rich. Web surveys, for example, are rich in paradata for measurement error evaluation, as Chapter 11 by Mario